DOCUMENT RESUME

ED 431 788 TM 029 873

AUTHOR Dimitrov, Dimiter M.
TITLE Mathematics and Science Achievement Profiles by Gender, Race, Ability, and Type of Item Response.
PUB DATE 1999-04-00
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Ability; *Academic Achievement; Cultural Differences; *Ethnic Groups; Grade 5; Intermediate Grades; *Mathematics Achievement; Mathematics Tests; Racial Differences; Responses; *Science Education; Science Tests; *Sex Differences; State Programs; *Test Items; Testing Programs
IDENTIFIERS Ohio; *Science Achievement

ABSTRACT

The purpose of this study was to determine patterns of gender and ethnic differences in science and mathematics achievement of fifth graders, taking into account student ability, item response format, and strands of learning outcomes. Subjects were 2,551 fifth graders from a large urban area in Ohio. The Ohio Off-Grade Proficiency Test was used. In mathematics, there were no gender differences, but there were ethnic differences that varied across ability levels and item response formats. In science, there were gender differences. The differences did not depend on ethnicity, but did vary across response formats, ability levels, and strands of learning outcomes. At the high ability level, boys did better than girls on open-ended format in physical sciences, but no gender differences were found in other science areas (nature of science, earth and space sciences, and life sciences). There were no gender differences in science achievement for the low and medium ability students. Regardless of gender, the largest gap between the achievement of the low, medium, and high ability students was on the open-ended format in physical sciences. (Contains 4 figures, 2 tables, and 34 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: TM

In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view. 9E

ED 431 788

Mathematics and Science Achievement Profiles by Gender, Race, Ability, and Type of Item Response

Dimiter M. Dimitrov

Kent State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Dimiter Dimitrov

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

TM029873

Paper presented at the Annual Meeting of the

American Educational Research Association

April 19-23, 1999, Montreal

BEST COPY AVAILABLE

Abstract

The purpose of this study was to determine patterns of gender and ethnic differences in science and mathematics achievement of fifth graders taking into account student ability, item response format, and strands of learning outcomes. In mathematics, there were no gender differences, but there were ethnic differences that varied across ability levels and item response formats. In science, there were gender differences. They did not depend on ethnicity, but did vary across response formats, ability levels, and strands of learning outcomes. At the high ability level, boys did better than girls on the open-ended format in physical sciences, but no gender differences were found in other science areas (nature of science, earth and space sciences, and life sciences). There were no gender differences in science achievement for the low and medium ability students. Regardless of gender, the largest gap between the achievement of the low, medium, and high ability students was on the open-ended format in physical sciences.

Mathematics and Science Achievement Profiles by

Gender, Race, Ability, and Type of Item Response

Issues of gender differences in science have received serious attention in the research in science education for the last two decades. Boys and girls have been compared on variables such as achievement, attitude, motivation, interest, and performance behaviors (e. g., Erickson & Erickson, 1984; Simpson & Oliver, 1985; Eccles & Blumenfield, 1985; Kahle et al., 1993; Greenfield, 1997; Morrell & Lederman, 1998; Jovanovich & King, 1998). Regarding gender differences in science achievement, the need for more detailed analysis is indicated in many previous studies (e. g., Erickson & Erickson 1984; Walford, 1980; Murphy, 1982; Saner et al., 1994; DeMars, 1998). It should be noted that most previous results about gender differences in science achievement are based on multiple-choice items and need to be revised in the light of the increased use of both multiple-choice and open-ended questions in many national and statewide assessment programs. The same is true for previous studies on gender and ethnic differences in mathematics achievement (e. g., Doolitle & Clearly, 1987; Cooper & Dorr, 1995). For both science and mathematics, very little is known about differential effects of student related factors (e.g., ethnicity, science ability) and test related factors (e. g., item format, learning outcomes) on gender differences in science achievement. When such effects are not taken into account, the results related to gender differences are of little value, if not misleading.

Purpose

The purpose of this study was to determine patterns of gender and ethnic differences in mathematics and science achievement across response formats and strands of learning outcomes, taking into account the role of student ability. Based on the results from a proficiency test for grade five, patterns of gender and ethnic differences were studied across two response formats (multiple-choice, open-ended) and three ability levels (low, medium, high). *Mathematics ability* and *science ability* are two different latent traits underlying the student performance on the

mathematics and science part, respectively, of the proficiency test used in this study. Using an appropriate item response theory model, the student ability scores in mathematics and science were determined from the responses on all items, including partial credit scores on open-ended items, of the mathematics and science proficiency tests. This provides more reliable estimation of the student ability compared to previous studies on gender differences in math and science using only multiple-choice scores to control for student ability level (e.g., DeMars, 1998).

Method

Subjects

The subjects for this study were 2551 fifth graders for a large urban area in North-East Ohio. Gender and ethnicity information was available for 2414students: 917 Caucasians (437 females, 480 males), 1167 African-American (588 females, 579 males), and 330 Hispanic (151 females, 179 males). It is expected that fifth grade is representative for a summative experience in science and mathematics education at the elementary school level.

Instrument

The Ohio Off-Grade Proficiency Test (OOPT) for grade five was used (Riverside Publishing, 1995). The mathematics part (OOPT-M) includes 30 multiple-choice and 10 open-ended items. The science part (OOPT-S) includes 32 multiple-choice and 10 open-ended items. The science test items are grouped by the publisher into four strands of learning outcomes: (a) Nature of Science, (b) Physical Sciences, (c) Earth and Space Sciences, and (d) Life Science. In this study, these strands were taken into account because previous research reported that boys outperform girls in Physics and not in other science areas. For this sample, the reliability of the mathematics ($\alpha = .83$) and science ($\alpha = .82$) test was adequate for group comparisons. The Pearson correlation between the multiple-choice and open-ended sections of the OOPT-M and OOPT-S was .70 and .50, respectively. This indicates that, in both OOPT-M and OOPT-S, the multiple-choice and open-ended items measure somewhat similar, but different in complexity and depth, constructs.

Procedures

In item response theory, the term "ability" connotes a latent trait that underlies the student's performance on a test (see, e. g., Hambleton et al., 1991, p. 77). The ability score of any student determines the probability for this student to answer correctly any test item. The units of the ability scale, called "logits", represent natural logarithms of odds ratios for success on the test items. In this study, the ability scores were calculated using the program PARSCALE which takes into account the partial credit scores of the students on the constructed-response items (Muraki & Bock, 1996; Muraki, 1992). Students with ability scores in the lower 27% were assigned to the low ability group, those with ability scores in the upper 27%, to the high ability group, and all other students, to the medium ability group. This procedure was chosen because the ability scores of all 2551 students were normally distributed on both OOPT-M and OOPT-S (Kelley, 1939). Multivariate analysis of variance (MANOVA) procedures were used with three between-subjects factors (Gender, Ethnicity, Ability) and a within-subject factors Format (multiple-choice, open-ended). The within-subject factor Strand (nature of science, physical sciences, earth and space sciences, life sciences) was used only for the OOPT-S. For each response format and strand of learning outcomes, the average scores of the students were presented on a common T-scale ($M = 50$, $SD = 10$).

Results

Mathematics

The MANOVA results for the OOPT-M showed no significant interactions between gender and the other factors: Gender x Race ($\Lambda = .997$, $F(4, 4790) = 1.85$, $p = .115$), Gender x Ability ($\Lambda = .999$, $F(4, 4790) = 0.69$, $p = .60$), and Gender x Race x Ability ($\Lambda = .996$, $F(8, 4790) = 0.99$, $p = .26$). For this reason, the differences between boys and girls across multiple-choice and open-ended response formats were studied regardless of ethnicity and ability level of the students. The results showed no significant gender effect ($\Lambda = .998$, $F(2, 2473) = 2.72$, $p = .066$). On the other

side, the interaction Race x Ability was significant ($\Lambda = .986$, $F(8, 4790) = 4.17$, $p < .0001$). That is why, the ethnic differences across the response formats were studied separately for the low, medium, and high ability students. The descriptive statistics (on a T-scale) are given in Table 1.

For the low ability students, the achievement profiles are given in Figure 1. There were no significant differences between the ethnic groups ($\Lambda = .986$, $F(4, 1304) = 2.30$, $p = .057$).

For the medium ability students, the achievement profiles are given in Figure 2. There were significant differences between the ethnic groups ($\Lambda = .984$, $F(4, 2236) = 4.40$, $p = .002$). More specifically, there were differences on the multiple-choice items ($F(2, 1119) = 8.68$, $p < .0001$) where Caucasians outperformed African Americans ($p = .002$) and Hispanics ($p = .001$), but there was no difference between African Americans and Hispanics ($p = .582$). There were no differences between the ethnic groups on the open-ended items ($F(2, 1119) = 0.65$, $p = .524$).

For the high ability students, the achievement profiles are given in Figure 3. There were significant differences between the ethnic groups ($\Lambda = .959$, $F(4, 1264) = 6.63$, $p < .0001$) on both the multiple-choice ($F(2, 633) = 8.74$, $p < .0001$) and open-ended ($F(2, 633) = 7.26$, $p = .001$) items. On the multiple-choice items, Caucasians performed better than African American ($p < .0001$) and Hispanics ($p = .030$), but there was no difference between African Americans and Hispanics ($p = .961$). On the open-ended items, Caucasians performed better than African Americans ($p < .0001$), but there were no other significant differences between the ethnic groups.

Science

The results from MANOVA showed a significant gender effect ($F(3, 2394) = 12.91$, $p < .01$), but no significant interaction between gender and ethnicity ($F(8, 4786) = 1.605$, $p = .118$). That is why, gender differences were analyzed regardless of ethnicity. Also, the significance of the interaction Gender x Format x Ability ($F(4, 4938) = 2.752$, $p < .05$) shows that the difference between boys and girls in science achievement depends on response formats and ability levels. For this reason, patterns of gender differences across the multiple-choice and open-ended items

of the strands of learning outcomes were studied separately for the low, medium, and high ability students. For each ability level, MANOVA was conducted for the significance test of differences between boys and girls across eight combinations of two response formats and four strands of learning outcomes. The omnibus MANOVA tests showed no significant gender differences for the low ($F(8, 654) = 1.818$, $p = .075$) and medium ($F(8, 1131) = 1.447$, $p = .173$) ability students. Significant gender differences were found for the high ability students ($F(8, 664) = 2.003$, $p < .05$). For this reason, the achievement profiles of the high ability students are given separately for boys and girls, whereas those for the low and medium ability students are combined for boys and girls (see Figure 4). The MANOVA tests of between-subjects effects for the high ability students showed that the difference between boys an girls was statistically significant only for the open-ended format in physical sciences ($F(1, 671) = 8.671$, $p < .01$). The descriptive statistics for the achievement profiles in Figure 4 are given in Table 2.

Discussion and Implications

The results from this study showed that gender did not play a significant role in the mathematics performance of fifth-graders on the Ohio Off-Grade Proficiency Test. This finding is consistent with those reported in previous studies related to gender differences in mathematics (e. g., Lewis & Hoover, 1986; DeMars, 1998; O'Neil & Brown, 1998). Statistically significant differences were found in the mathematics performance of Caucasian, African American, and Hispanic students. However, the differences between the three ethnic groups across the low, medium, and high ability level did not consistently favor one group over others. Also, the Wilk's lambda (Λ) in the MANOVA tests for ethnic differences on the OOPT-M varied from .984 to .959 which indicates small effect size in these differences.

The comparison of Caucasian and Hispanic students supports the results of O'Neil and Brown (1998) indicating that, for eight-graders from California schools, Caucasians performed better than Hispanics on the multiple-choice items, but there was no difference between these

two ethnic groups on the open-ended items. The results also showed that Caucasians did not consistently outperform African Americans in mathematics. For example, the low and medium ability students from these two ethnic groups did not differ on the open-ended items. The third comparison, between African Americans and Hispanics, also varied across ability levels. High ability Hispanics did better than high ability African Americans, but there were no differences between these two ethnic groups at the low and medium ability levels.

Although the OOPT-M multiple-choice and open-ended items measure positively correlated characteristics of students math ability, the results show that the response formats have differential effect on student performance. It seems that the low and high ability Hispanics were most affected by the response format in their math achievement. They performed better on the open-ended than on the multiple-choice items. This suggests that (a) the math performance of the Hispanic students was not negatively influenced by their level of proficiency with the English language, and (b) Hispanic students may need help improving their response strategies on the multiple-choice items.

In science, there were gender differences and they did not depend on ethnicity. However, item response formats, ability level, and strands of learning outcomes have differential effects on student performance. There were no gender differences for the low and medium ability students, regardless of response formats and strands of learning outcomes. For the high ability students, boys outperformed girls on the open-ended format for physical sciences and there were no other significant gender differences. The examination of Figure 4 shows that the high ability girls performed almost equally well on the multiple-choice and open-ended items. High ability boys, however, performed better on open-ended than on multiple-choice items, especially in physical sciences. Regardless of gender, the open-ended items in physical sciences were the easiest strand for the high ability students and the most difficult one for the low and medium ability students. For the multiple-choice items, the achievement profiles of the students from all ability levels

were very close and stable across the four strands of learning outcomes. It should be noted that interaction effects of response formats and ability levels can not be appropriately identified when the partial credit scores on the open-ended items are not taken into account in determining ability scores.

Previous studies also reported that boys do better than girls in the physical sciences and the gender differences are not substantial in other subject areas (e. g., Erickson & Erickson, 1984; NAEP, 1979), but they do not take into account differential effects of ability levels and response formats. According to the sociological interpretation of some authors, the advantage of boys in physical sciences is mostly due to their (a) previous experience from hobbies and games, and (b) greater motivation, interest, and positive attitude toward science fostered by sex-typed beliefs that science is still a male domain (see, e. g., Erickson & Erickson, 1984; Johnson, 1987; Kelly, 1988; Erickson & Farkas, 1991; Jovanovich & King, 1998).

One general explanation of the differential effect of item response formats comes from previous research indicating that students employ different lines of reasoning in dealing with multiple-choice and open-ended questions (e. g., Frederiksen, 1994; O'Neil & Brown, 1998). Further development of this explanation is needed because, as indicated by the results in this study, the differential effect of response formats may vary across ability levels and strands of learning outcomes. This, along with factors such as test content and age differences, may explain discrepancies between the findings in this study and those reported by DeMars (1998), for eight graders, that boys outperform girls on the multiple-choice items and girls do as well as or better than boys on the constructed- response items (p. 292).

Knowledge about the patterns of gender differences revealed in this study is important for informed teaching strategies and decisions related to the science education of fifth graders. The results support implications from previous research about the necessity of diagnostic assessment and treatment (e. g., hands-on science lessons) that can compensate for disparities in boys' and

girls' science-related experiences outside of school (e. g., Erickson & Erickson, 1984; Kahle et al.,1993; Greenfield, 1997; Jovanovich & King, 1998). Particular attention must be attributed to research, activity, and intervention in science education that may contribute to reducing the gaps in open-ended physics questions between the low, medium, and high ability students.

In conclusion, regarding both mathematics and science, the results in this study provide evidence against the generality of perceptions such as (a) students perform lower on open-ended items, (b) multiple-choice items capture lower level skills, or (c) boys do better than girls in math and science. Awareness of the differential effects of factors such as gender, ability, item response format, and learning outcomes on student achievement is a valuable source for mathematics and science teachers who want their instruction to be effective and responsive to proficiency testing.

References

Allen, M. J., & Yen, W. M. (1979). Introduction to Measurement Theory. Monterey, CA: Brooks/Cole.

Cooper, H., & Dorr, N. (1995). Race comparison on need for achievement: a meta-analytic alternative to Graham's narrative review. Review of Educational Research, 65, 483-508.

DeMars, C. E. (1998). Gender differences in mathematics and science on a high school proficiency exam: The role of response format. Applied Measurement in Education, 11, 279-299.

Doolittle, A. E., & Clearly, T. A. (1987). Gender-based differential item performance in mathematics achievement items. Journal of Educational Measurement, 24, 157-166.

Eccles, J. S., Blumenfeld, P. (1985). Classroom experiences and student gender: Are there differences and do they matter? In L. C. Wilkinson & C. B. Marrett (Eds.), Gender influences in classroom interaction (pp. 79-114). New York: Academic Press.

Erickson, G. A., Erickson, L. J. (1984) Achievement: evidence, explanations, and implications. Science Education, 68, 63-89.

Erickson, G., & Farkas, S. (1991). Prior experience and gender differences in science achievement. The Alberta Journal of Educational Research, 37, 225-239.

Frederiksen, N. (1984). The real test bias: influences of testing on teaching and learning. American Psychologist, 39, 193-202.

Gray, J A. (1981). A biological basis for the sex differences in achievement in science? In A. Kelly (Ed.), The missing half: girls and science education. Manchester: Manchester Univ. Press.

Greenfield, T. A. (1997). Gender- and grade-level differences in science interest and participation. Science Education, 81, 259-276.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of items response theory. Newbury Park, CA: SAGE.

Johnson, S. (1987). Gender differences in science: Parallels in interest, experience, and

performance. International Journal of Science Education, 9, 467-481.

Jovanovich, J., King, S. S. (1998). Boys and girls in the performance-based science classroom: who's doing the performing? American Educational Research Journal, 35, 477- 496.

Kahle, J. B., & Lakes, M. (1983). The myth of equality in science classrooms. Journal of Research in Science Teaching, 20, 131-140.

Kahle, J. B., Parker, L. H., Rennie, L. J., & Riley, D. (1993). Gender differences in science education: Building a model. Educational Psychologist, 28, 379-404.

Kelly, A. (1988). Sex stereotypes and school science. A three-year follow-up. Educational Studies, 14, 151-163.

Kelly, T. L. (1939). The selection of upper and lower groups for the validation of test items. Journal of Educational Psychology, 30, 17-24. Monterey, CA: Brooks/Cole.

Lewis, J. C., & Hoover, H. D. (1983, April). Sex differences on standardized academic achievement tests. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Meyr, M., & Koehler, M. S. (1990). Internal influences on gender differences in mathematics. In E. Fennema & G. Leder (Eds.), Mathematics and gender (pp. 60-95). New York: Teachers College Press.

Morrell, P. D., & Lederman, N. G. (1998). Student's attitudes toward school and classroom science: Are they independent phenomena? School Science and Mathematics, 98, 76-82.

Muraki, E. (1992). A generalized partial credit model: Applications of an EM algorithm. Applied Psychological Measurement, 16, 159-176.

Muraki, E., & Bock, R. D. (1996). PARSCALE: IRT based test scoring and item analysis for graded open-ended exercises and performance tasks [Computer program]. Chicago, IL: Scientific Software Inc.

Murphy, R. J. L. (1982). Sex differences in objective test performance. British Journal of

Educational Psychology, 52, 213-219.

National Assessment of Educational Progress (NAEP). Three assessments of science 1969-1977: technical summary. Denver: Education Commissions of the States, 1979.

O'Neil, H. F., & Brown, R. S. (1998). Differential effects of question formats in math assessment on metacognition and affect. Applied Measurement in Education, 11, 331-351.

Riverside Publishing (1997). Ohio Off-Grade Proficiency Tests - Mathematics.; Science.. Chicago, IL: Author.

Saner, H., Klein, S., Bell, R., & Comfort, K. (1994). The utility of multiple raters and tasks in science performance assessments. Educational Assessment, 2, 257-272.

Sells, L. (1976). Images of science and scientists and science education. Science Education, 76, 139-149.

Simpson, R. D., & Oliver, J. S. (1985). Attitude toward science and achievement motivation profiles of male and female science students in grades six through ten. Science Education, 69, 511-526.

SPSS Inc. (1997). SPSS (Windows version 7.5): User's guide. Chicago, IL; Author

Steinkamp, M. W., & Maehr, M. L. (1983). Affect, ability, and science achievement; A quantitative synthesis of correlational research. Review of Educational Research, 53, 369-396.

Walford, G. (1980). Sex bias in physics textbooks. School Science Review, 62, 220-227.

Weinburgh, M. (1995). Gender differences in student attitudes toward science: A meta-analysis of the literature 1970 - 1991. Journal of Research in Science Teaching, 32, 387-398.

Wightman, L. F. (1998). An examination of sex differences in LSAT scores from the perspective of social consequences. Applied Measurement in Education, 11, 255-277.

Table 1

OOPT-M Descriptive Statistics by Ability, Ethnicity, and Response Format

	Ability Level								
	Low			Medium			High		
Ethnic group:	C	AA	H	C	AA	H	C	AA	H
Response Format									
Multiple-Choice									
M	39.77	39.86	38.34	50.82	49.38	48.78	61.65	59.61	59.20
SD	6.21	5.70	5.25	6.70	5.99	5.61	6.39	6.34	7.17
n	196	343	117	411	567	144	310	257	69
Open-Ended									
M	40.99	41.33	41.51	48.33	48.62	48.64	62.12	60.09	61.72
SD	2.55	2.75	2.65	4.15	4.23	4.36	6.63	5.96	6.95
n	196	343	117	411	567	144	310	257	69

Note. C = Caucasians, AA = African American, H = Hispanic. All scores are presented on a common T-scale (M = 50, SD = 10).

Table 2
OOPT-S Descriptive Statistics for High Ability Boys, High Ability Girls, Medium Ability (Boys and Girls), and Low Ability (Boys and Girls)

	Ability Level							
	High				Medium (Boys and Girls)		Low (Boys and Girls)	
	Boys (n = 372)		Girls (n = 301)		n = 1140		n = 663	
Format / Strand	M	SD	M	SD	M	SD	M	SD
Multiple-Choice								
Physics	54.68	1.75	54.48	1.78	53.45	1.66	52.40	1.54
NS	55.23	1.62	55.12	1.62	53.89	1.38	52.83	1.31
E & S	54.86	1.51	54.92	1.50	53.87	1.55	52.58	1.30
LS	54.94	2.36	55.24	2.07	54.08	2.42	52.70	2.15
Open-Ended								
Physics	62.08	9.09	59.98	9.28	47.95	6.76	42.34	4.02
NS	57.36	4.88	57.18	4.99	53.52	3.87	51.06	2.20
E & S	53.71	3.99	53.73	4.22	51.26	2.07	50.47	1.22
LS	55.87	12.07	55.86	11.16	49.25	9.15	45.45	5.82

Note. NS = Nature of Science; E & S = Earth and Space Sciences; LS = Life Sciences. All scores are represented on a common T-scale (M = 50, SD = 10).

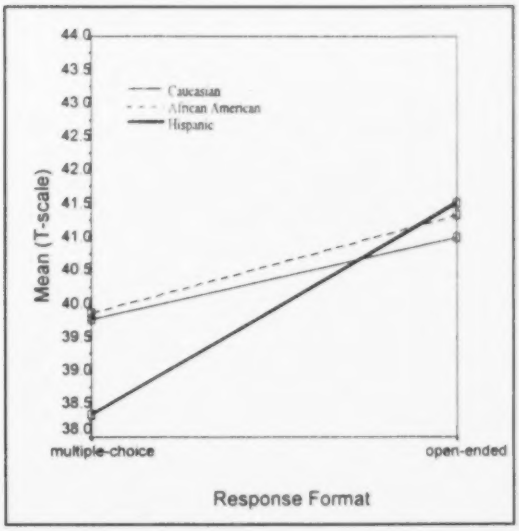

Figure 1. Mathematics achievement profiles for the low ability students

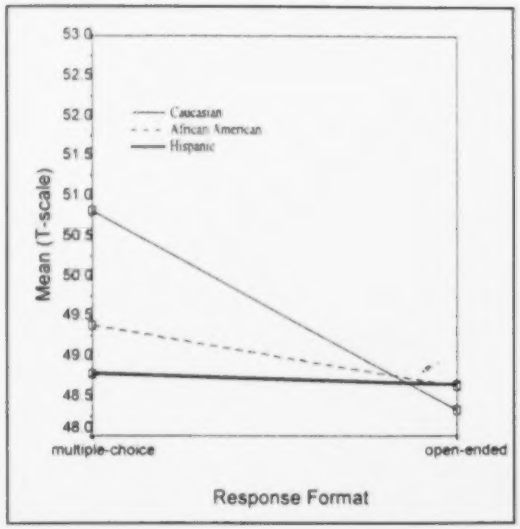

Figure 2. Mathematics achievement profiles for the medium ability students

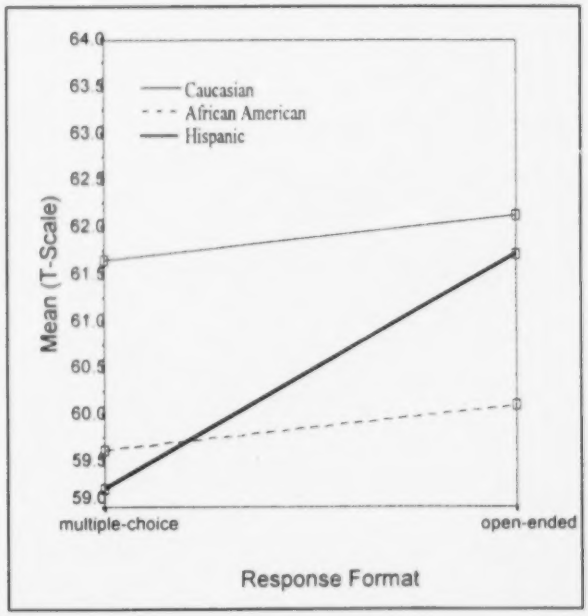

Figure 3. Mathematics achievement profiles for the

BEST COPY AVAILABLE

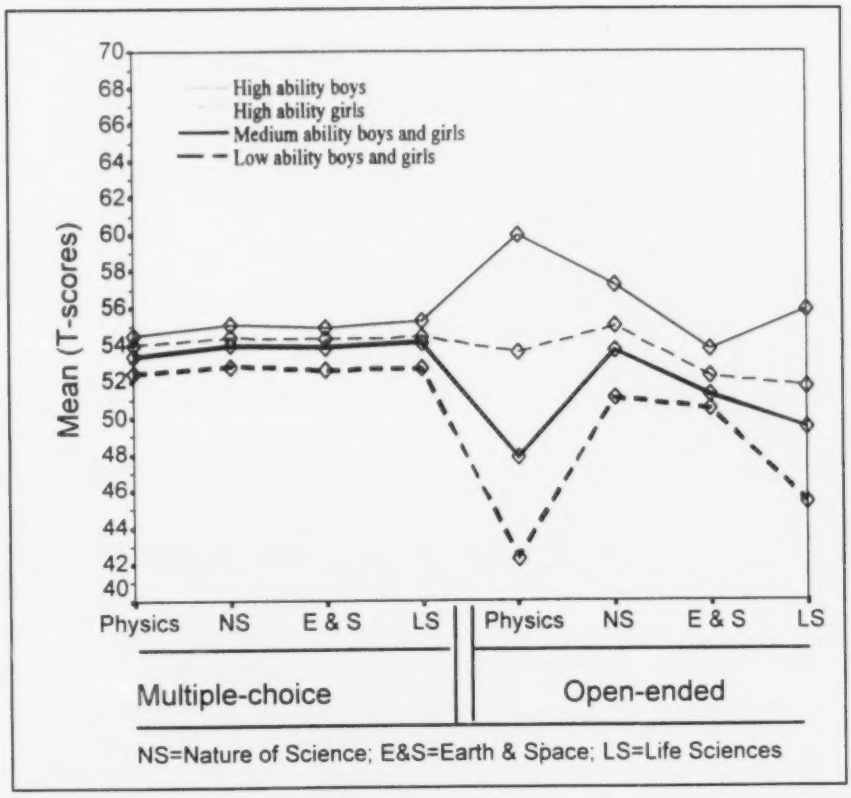

Figure 4. Science achievement profiles across strands of learning outcomes, by gender and ability level.

BEST COPY AVAILABLE